

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.

2455-4602

Total Pages

First Named Inventor or Application Identifier

Mazin RAHIM et al.

Express Mail Label No.

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

1. ☒ Fee Transmittal Form
(submit an original, and a duplicate for fee processing)
2. ☒ Specification [Total Pages 13]
(preferred arrangement set forth below)
3. ☒ Drawing(s) (35 USC 113) [Total Sheets 2]
4. ☒ Oath or Declaration [Total Pages 2]
- a. ☒ Newly executed (original or copy)
- b. ☐ Copy from a prior application (37 CFR 1.63(d))
(for continuation/divisional with Box 17 completed)
[Note Box 5 below]
- i. ☐ DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s)
named in the prior application, see 37 CFR
163(d)(2) and 1.33(b)
5. ☐ Incorporation by reference (useable if Box 4b is checked)
The entire disclosure of the prior application, from which a copy of the oath
or declaration is supplied under Box 4b, is considered as being part of the
disclosure of the accompanying application and is hereby incorporated by
reference herein.

ADDRESS TO:

Assistant Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

6. ☐ Microfiche Computer Program (Appendix)
7. ☐ Nucleotide and/or Amino Acid Sequence Submission
(if applicable, all necessary)
- a. ☐ Computer Readable Copy
- b. ☐ Paper Copy (identical to computer copy)
- c. ☐ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

8. ☒ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney
10. ☐ English Translation Document (if applicable)
11. ☐ Information Disclosure ☐ Copies of IDS
Statement (IDS)/PTO-1449 Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)
14. ☐ Small Entity ☐ Statement filed in prior application,
Statement(s) Status still proper and desired
15. ☐ Certified Copy of Priority Document(s)
(if foreign priority is claimed)
16. ☐ Other :

17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior Application No:

18. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label

(Insert Customer No. or Attach bar code label here)

or ☒ Correspondence address below

NAME	Christopher J. Hamaty, Esquire				
ADDRESS	MORGAN & FINNEGAN, L.L.P., 345 Park Avenue				
CITY	New York	STATE	New York	ZIP CODE	10154
COUNTRY	United States of America			FAX	202-857-7929

19. SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED

NAME	Christopher J. Hamaty		Reg. #	37634
TELEPHONE	202-857-7887			
SIGNATURE	<i>Chris Hamaty</i>		DATE	5-19-99

"Express Mail" Mailing Label Number

Date of Deposit

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington D.C., 20231

The PTO did not receive the following
listed item(s) *Assignment*

(Printed Name of Person Mailing Paper)

(Signature of Person Mailing Paper)

RECOGNIZING THE NUMERIC LANGUAGE IN NATURAL SPOKEN DIALOGUE

BACKGROUND OF THE INVENTION

5 1. Field of the Invention

This invention relates to a system for numeric language recognition in natural spoken dialogue.

2. Description of the Related Art

10 Speech recognition is a process by which an unknown speech utterance (usually in the form of a digital PCM signal) is identified. Generally, speech recognition is performed by comparing the features of an unknown utterance to the features of known words or word strings. Hidden Markov models (HMMs) for automatic speech recognition (ASR) rely on high dimensional feature vectors to summarize the short-time, acoustic properties of speech. 15 Though front-ends vary from speech recognizer to speech recognizer, the spectral information in each frame of speech is typically codified in a feature vector with thirty or more dimensions. In most systems, these vectors are conditionally modeled by mixtures of Gaussian probability density functions (PDFs).

20 Recognizing connected digits in a natural spoken dialog plays a vital role in many applications of speech recognition over the telephone. Digits are the basis for credit card and account number validation, phone dialing, menu navigation, etc.

Progress in connected digit recognition has been remarkable over the past decade. For databases recorded under carefully monitored laboratory conditions, speech recognizers have been able to achieve less than 0.3% word error rate. Dealing with telephone 25 speech has added a new dimension to this problem. Variations in the spectral characteristics due to different channel conditions, speaker populations, background noise and transducer

equipment cause a significant degradation in recognition performance. Previous practice has strictly focused on dealing with constrained input speech to produce digit sequences.

SUMMARY OF THE INVENTION

5 In accordance with the principles of the invention, the set of words or phrases that are relevant to the task of understanding and interpreting number strings is referred to as the "numeric language". The "numeric language" defines the set of words or phrases that play a key role in the understanding and automation of users' requests. According to an exemplary embodiment of the invention, the numeric language consists of the set of word or phrase
10 classes that are relevant to the task of understanding and interpreting number strings, such as credit card numbers, telephone numbers, zip codes, etc., and consists of six distinct phrase classes including "digits", "natural numbers", "alphabets", "restarts", "city/country name", and "miscellaneous".

15 In the exemplary embodiment of the invention, a system includes a speech recognition processor that receives unconstrained fluent input speech and produces a string of words that can include a numeric language, and a numeric understanding processor that converts the string of words into a sequence of digits based on a set of rules. An acoustic model database utilized by the speech recognition processor includes a first set of hidden Markov models that characterize the acoustic features of numeric words, a second set of hidden
20 Markov models that characterize the acoustic features of the remaining vocabulary words, and a filler model that characterizes the acoustic features of out-of-vocabulary utterances. An utterance verification processor verifies the accuracy of the string of words. A validation database stores a grammar, and a string validation processor outputs validity information based on a comparison of the sequence of digits with the grammar. A dialogue manager processor
25 initiates an action based on the validity information.

Other aspects and advantages of the invention will become apparent from the following detailed description and accompanying drawing, illustrating by way of example the features of the invention.

5

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a numeric language recognition system in accordance with the principles of the invention; and

FIG. 2 illustrates an acoustic model database in accordance with the principles of the invention.

DETAILED DESCRIPTION

For a better understanding of the invention, together with other and further objects, advantages, and capabilities thereof, reference is made to the following disclosure and the figures of the drawing. For clarity of explanation, the illustrative embodiments of the present invention are presented as comprising individual functional blocks. The functions these blocks represent may be provided through the use of either shared or dedicated hardware, including, but not limited to, hardware capable of executing software. For example, the functions of the blocks presented in FIG. 1 may be provided by a single shared processor. Illustrative embodiments may comprise digital signal processor (DSP) hardware, read-only memory (ROM) for storing software performing the operations discussed below, and random-access memory (RAM) for storing DSP results. Very large scale integration (VLSI) hardware embodiments, as well as custom VLSI circuitry in combination with a general purpose DSP circuit, may also be provided. Use of DSPs is advantageous since the signals processed represent real physical signals, processes and activities, such as speech signals, room background noise, etc.

This invention is directed to advancing and improving numeric language recognition in the telecommunications environment, particularly the task of recognizing numeric

words when embedded in natural spoken dialog. In particular, the invention is directed toward the task of recognizing and understanding users' responses when prompted to respond with information needed by an application involving the numeric language, such as, for example, their credit card or telephone number. We have identified those words that are relevant to the task and enhance the performance of the system to recognize those relevant words.

By way of example, and not limitation, in a specific embodiment of the invention, the numeric language forms the basis for recognizing and understanding a credit card and a telephone number in fluent and unconstrained spoken input. Our previous experiments have shown that considering the problem of recognizing digits in a spoken dialogue as a large-vocabulary continuous speech recognition task, as opposed to the conventional detection methods, can lead to improved system performance.

In an exemplary system for recognizing the numeric language in a natural spoken dialogue, illustrated in FIG. 1, a feature extraction processor 12 receives input speech. A speech recognition processor 14 is coupled to the feature extraction processor 12. A language model database 16 is coupled to the speech recognition processor 14. An acoustic model database 18 is coupled to the speech recognition processor 14.

A numeric understanding processor 20 is coupled to the speech recognition processor 14. An utterance verification processor 22 is coupled to the speech recognition processor 14. The utterance verification processor 22 is coupled to the numeric understanding processor 20. The utterance verification processor 22 is coupled to the acoustic model database 18.

A string validation processor 26 is coupled to the numeric understanding processor 20. A database 28 for use by the string validation processor 26 is coupled to the string validation processor 26.

A dialog manager processor 30 is coupled to the string validation processor 26. The dialogue manager processor 30 initiates action according to the invention in response to the results of the string validation performed by the string validation processor 26.

Using a spoken dialogue system imposes a new set of challenges in recognizing
5 digits, particularly when dealing with naive users of the technology. In this example, during a spoken dialogue users are prompted with various open questions such as, "What number would you like to call?", "May I have your card number please?", etc. The difficulty in automatically recognizing responses to such open questions is not only to deal with fluent and unconstrained speech, but also to be able to accurately recognize an entire string of numerics (i.e., digits or
10 words identifying digits) and/or alphabets. In addition the system ought to demonstrate robustness towards out-of-vocabulary words, hesitation, false-starts and various other acoustic and language variabilities.

Performance of the system was examined in a number of field trial studies with customers responding to the open-ended prompt "How may I help you?" with the goal to provide
15 an automated operator service. The purpose of this service is to recognize and understand customers' requests whether it relates to billing, credit, call automation, etc.

In an important part of the field trials, customers were prompted to say a credit card number or a telephone number to obtain call automation or billing credit. Various types of prompts were studied with the objective to stimulate maximally consistent and informative
20 responses from large populations of naive users. These prompts are engineered towards asking users to say or repeat their credit card or telephone number without imposing rigid format constraints.

The system is optimized to recognize and understand words in the dialogue that are salient to the task. Salient phrases are essential for interpreting fluent speech. They are
25 commonly identified by exploiting the mapping from unconstrained input to machine action.

with its frame log energy are augmented with their first and second order time derivatives. The energy coefficient, normalized at the operating look-ahead delay, is also applied for end-pointing the speech signal.

Accurate numeric recognition in fluent and unconstrained speech clearly
5 demands detailed acoustic modeling of the numeric language (the numeric words and phrases). It is essential to accurately model out-of-vocabulary words (the non-numerics) as they constitute over eleven percent of the database. Accordingly, our design strategy for the acoustic model 18 has been to use two sets of subword units. Referring to FIG. 2, a first set 36 of hidden Markov models (HMMs) that characterize the acoustic features of numeric words is dedicated for the
10 numeric language. A second set 38 of HMMs that characterize the acoustic features of the remaining vocabulary words is dedicated for the remaining vocabulary words. Each set 36, 38 applies left-to-right continuous density hidden Markov models (HMMs) with no skip states.

In the first set 36 dedicated for recognition of numerics, context-dependent
acoustic units have been used which captured all possible inter-numeric coarticulation. The
15 basic structure is that each word is modeled by three segments; a head, a body and a tail. A word generally has one body, which has relatively stable acoustic characteristics, and multiple heads and tails depending on the preceding and following context. Thus, junctures between numerics are explicitly modeled. Since this results in a huge number of subword units, and due to the limited amount of training data, the head-body-tail design was strictly applied for the
20 eleven digits (i.e., "one", "two", "three", "four", "five", "six", "seven", "eight", "nine", "zero", and "oh"). This generated two hundred seventy-four units which were assigned a three-four-three state topology corresponding to the head-body-tail units, respectively.

The second set 38 of units includes forty tri-state context-independent subwords
that are used for modeling the non-numeric words, which are the remaining words in the
25 vocabulary. Therefore, in contrast to traditional methods for digit recognition, out-of-vocabulary

words are explicitly modeled by a dedicated set of subword units, rather than being treated as filler phrases.

To model transitional events between numerics, non-numerics and background/silence, an additional set 40 of units is used. Three filler models with different state topologies are also used to accommodate for extraneous speech and background noise events. In total, three hundred thirty-three units are employed in the exemplary units. Each state includes thirty-two Gaussian components with the exception of the background/silence model which includes sixty-four Gaussian components. A unit duration model, approximated by a gamma distribution, is also used to increment the log likelihood scores.

The language model database 16 is used by the speech recognition processor 14 to improve recognition performance. The language model database 16 contains data that describes the structure and sequence of words and phrases in a particular language. In this specific example, the data stored in the language model database 16 might indicate that a number is likely to follow the phrase "area code" or that the word "code" is likely to follow the word "area"; or, more generally, the data can indicate that in the English language, adjectives precede nouns, or in the French language, adjectives follow nouns. While language modeling is known, the combination of the language model database 16 with the other components of the system illustrated in FIG. 1 is not known.

Speech, or language, understanding is an essential component in the design of spoken dialogue systems. The numeric understanding processor 20 provides a link between the speech recognition processor 14 and the dialogue manager processor 30 and is responsible for converting the recognition output into a meaningful query.

The numeric understanding processor 20 translates the output of the recognizer 14 into a "valid" string of digits. However, in the event of an ambiguous request or poor recognition performance, the numeric understanding processor 20 can provide several

hypotheses to the dialogue manager processor 30 for repair, disambiguation, or perhaps clarification.

A rule-based strategy for numeric understanding is implemented in the numeric understanding processor 20 to translate recognition results (e.g., N-best hypotheses) into a simplified finite state machine of digits only. Several classes of these rules which aim to translate input text into a digit sequence are presented in TABLE 1.

Rule	Definition	Example
Naturals	translating natural numbers	one eight hundred and two → 1 8 0 0 2
Restarts	correcting input text	nine zero eight sorry nine one eight → 9 1 8
Alphabets	translating characters	A Y one two three → 2 9 1 2 3
City/Country	translating city/country area codes	calling London, England → 4 4 1 8 8
Numeric Phrases	realigning digits	nine on two area code nine zero one → 9 0 1 9 1 2
Out-of vocabulary	filtering	what is the code for Florham Park → 9 7 3

TABLE 1

The utterance verification processor 22 identifies out-of-vocabulary utterances and utterances that are poorly recognized. The utterance verification processor 22 provides the dialogue manager 30 with a verification measure of confidence that may be used for call confirmation, repair or disambiguation. The output of the utterance verification processor 22 can be used by the numeric understanding processor 20.

Information is validated before being sent to the dialogue manager processor 30. Due to ambiguous speech inputs and possible errors in the dialogue flow, sometimes

WHAT IS CLAIMED IS:

- 1 1. A system, comprising:
 - 2 a speech recognition processor that receives unconstrained input speech and
 - 3 outputs a string of words that can include a numeric language; and
 - 4 a numeric understanding processor that converts the string of words into a
 - 5 sequence of digits.

- 1 2. The system of claim 1, further comprising:
 - 2 an acoustic model database utilized by the speech recognition processor.

- 1 3. The system of claim 2, wherein the acoustic model comprises:
 - 2 a first set of hidden Markov models that characterize the numeric language; and
 - 3 a second set of hidden Markov models that characterize the remaining language
 - 4 in the vocabulary.

- 1 4. The system of claim 3, further comprising:
 - 2 a set of filler models that characterizes out-of-vocabulary features.

- 1 5. The system of claim 1, further comprising:
 - 2 an utterance verification processor that verifies the accuracy of the numeric
 - 3 language in the string of words.

1 6. The system of claim 1, further comprising:
2 a validity database that stores a grammar; and
3 a string validation processor that outputs validity information based on a
4 comparison of the sequence of digits with the grammar.

1 7. The system of claim 6, further comprising:
2 a dialogue manager processor that initiates an action based on the validity
3 information.

1 8. The system of claim 1, further comprising:
2 a language model database that emphasizes the numeric language utilized by
3 the speech recognition processor.

1 9. The system of claim 1, wherein:
2 the numeric understanding processor converts the string of words into the
3 sequence of digits based on a set of rules.

1 10. A method, comprising the steps of:
2 receiving unconstrained input speech and outputting a string of words that can
3 include a numeric language; and
4 converting the string of words into a sequence of digits.

666750 "E 947E 60

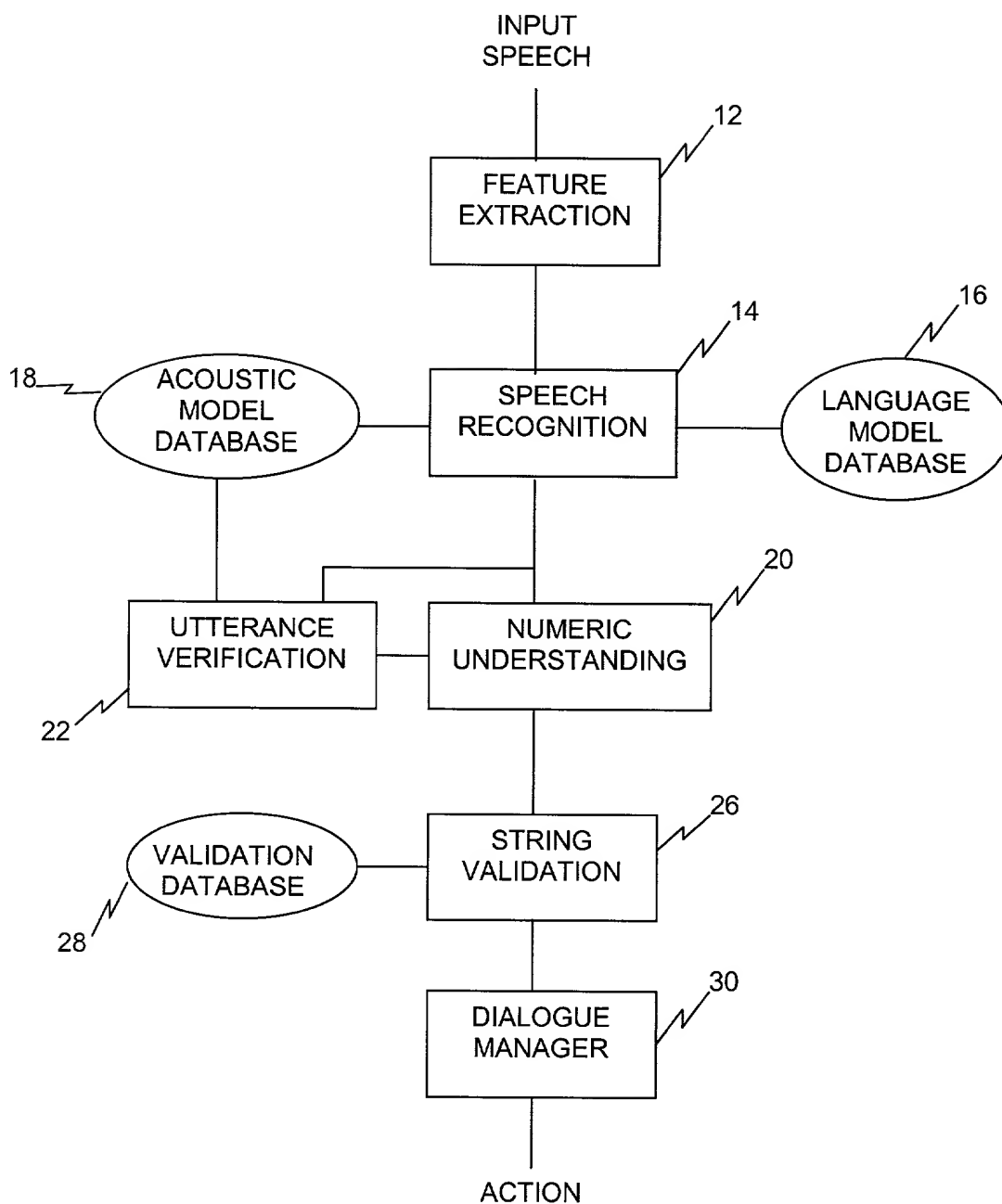


FIG. 1

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE

Declaration and Power of Attorney

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled "RECOGNIZING THE NUMERIC LANGUAGE IN NATURAL SPOKEN DIALOGUE" the specification of which is attached hereto.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by an amendment, if any, specifically referred to in this oath or declaration.

I acknowledge the duty to disclose all information known to me which is material to patentability as defined in Title 37, Code of Federal Regulations, 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

None

I hereby claim the benefit under Title 35, United States Code, 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, 112, I acknowledge the duty to disclose all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

None

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any

patent issued thereon.

I hereby appoint the following attorney(s) with full power of substitution and revocation, to prosecute said application, to make alterations and amendments therein, to receive the patent, and to transact all business in the Patent and Trademark Office connected therewith:

Alfred G. Steinmetz	(Reg. No. 22,971)
Samuel H. Dworetsky	(Reg. No. 27,873)
Robert Levy	(Reg. No. 28,234)
Thomas A. Restaino	(Reg. No. 33,444)
Jose A. De La Rosa	(Reg. No. 34,810)
Michele L. Conover	(Reg. No. 34,962)
Christopher A. Hughes	(Reg. No. 26,914)
Christopher J. Hamaty	(Reg. No. 37,634)

Please address all correspondence to: **Christopher J. Hamaty, Morgan & Finnegan, 345 Park Avenue, New York, NY 10154.** Telephone calls should be made to: **Christopher J. Hamaty** by dialing Area Code (202) 857-7887 .

Full name of 1st joint inventor: **Mazin G. Rahim**

Inventor's signature  Date 5/18/99

Residence: **1 Remington Court, Matawan, Monmouth County, New Jersey 07747, U.S.A.**

Citizenship: **U.S.A.**

Post Office Address: **1 Remington Court, Matawan, New Jersey 07747, U.S.A.**

Full name of 2nd joint inventor: **Giuseppe Riccardi**

Inventor's signature  Date 5/18/99

Residence: **716 Hudson Street, Apt. 1C, Hoboken, Hudson County, New Jersey 07030, U.S.A.**

Citizenship: **Italy**

Post Office Address: **716 Hudson Street, Apt. 1C, Hoboken, New Jersey 07030, U.S.A.**

SCANNED 1

Full name of 3rd joint inventor: **Jeremy Huntley Wright**

Inventor's signature _____ Date

Residence: **87 King George Road, Warren, Somerset County, New Jersey 07059, U.S.A.**

Citizenship: **United Kingdom**

Post Office Address: **87 King George Road, Warren, New Jersey 07059, U.S.A.**

Full name of 4th joint inventor: **Bruce Melvin Buntschuh**

Inventor's signature _____ Date

Residence: **10 Riverbend Road, Berkley Heights, Union County, New Jersey 07922, U.S.A.**

Citizenship: **U.S.A.**

Post Office Address: **10 Riverbend Road, Berkley Heights, New Jersey 07922, U.S.A.**

Full name of 5th joint inventor: **Allen Louis Gorin**

Inventor's signature Allen Louis Gorin Date 5/18/99

Residence: **205 Spring Ridge Drive, Berkeley Heights, Union County, New Jersey 07922, U.S.A.**

Citizenship: **U.S.A.**

Post Office Address: **205 Spring Ridge Drive, Berkeley Heights, New Jersey 07922, U.S.A.**